УДК 519.1

Н.Д. КОШЕВОЙ, Е.М. КОСТЕНКО, Н.В. ДОЦЕНКО, А.В. ПАВЛИК

Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Украина

МЕТОД ПЕРЕЧИСЛЕНИЯ СИМВОЛЬНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Рассмотрены свойства символьных последовательностей и способы их формального представления в виде серийных последовательностей. Введены операции на множестве символьных последовательностей, позволяющие проводить их преобразования. Исследованы способы описания серийных последовательностей и их типовых представителей. Предложен метод перечисления символьных последовательностей, позволяющий создавать каталоги типовых символьных последовательностей с заданными свойствами. Получены оценки количества типовых серийных последовательностей и процедура формирования множества типовых вариантов. Рассмотрены примеры реализации предложенного метода.

Ключевые слова: *символьная последовательность, серия, перечисление, типовой представитель.*

Постановка проблемы

Символьные последовательности являются классическим объектом математики, и методы их анализа развивались как в прикладных, так и в теоретических исследованиях (теория управления, теоретическое программирование, лингвистика, биология и др.). Как объект прикладного исследования символьные последовательности возникают во всех областях, где рассматриваются те или иные объекты, состоящие из большого числа одинаковых фрагментов [1]. Среди актуальных вопросов является перечисление типовых символьных последовательностей, позволяющее решить задачи типизации и унификации для широкого класса задач.

Анализ последних исследований и публикаций

Интенсивные исследования свойств символьных последовательностей связаны с развитием передовых систем канального кодирования. При этом символьную последовательность рассматривают как серийную последовательность, т. е. состоящую из последовательно расположенных серий символов. Символы, входящие в одну серию одинаковые. Основным направлением исследований стали коды с ограничениями на длины серий или канальные (RLL) коды. Развитие теории канального кодирования нашло отражение в работах К. Имминка, В.Брауна, [2], О.Ф. Курмаева [3]. Была предложена нумерационная процедура кодирования для канала с ограничениями, получены рекуррентные уравнения и производящие функции для перечисления последовательностей с ограничениями, что дало возможность изучать асимптотику и получать оценки для канала. Разработка и исследование кодовой конструкции для последовательностей с ограничениями на длины серий, вес или заряд рассмотрены в работе Курмаева [3].

В большинстве работ рассмотрены частные виды серийных последовательностей. В работах В.А. Амелькина [4, 5] рассматриваются множества целочисленных серийных последовательностей длины m, структура которых определяется ограничениями на количество, суммарную длину и допустимые длины натуральных серий [4], а в работе [5] приведена нумерация неубывающих и невозрастающих серийных последовательностей. Плотно вложенные последовательности исследованы в работе Д. Голица [6].

Анализ известных публикаций показывает, что в настоящее время исследованы свойства отдельных видов символьных последовательностей, в то время как символьные последовательности могут найти применение при решении многих задач: оптимальное планирование многофакторного эксперимента, техническая диагностика, управление проектами и др. Для этого необходимо исследовать свойства символьных последовательностей и разработать методы их перечисления.

Цель работы

Разработка метода перечисления символьных последовательностей.

Основные результаты исследований

Пусть имеется алфавит $\Omega = \{\omega_1, \ldots, \omega_r\}$, состоящий из r символов. Символьной (r, n) последовательностью назовем последовательность (слово)

$P = \{p_1,\ldots, p_n)$, в которой $p_i \in \Omega$, $i = 1,\ldots, n$; $n \ge r$ и в последовательности P представлены все символы из алфавита Ω.

Подпоследовательность $p_{t+1}p_{t+2}...p_{t+v}$ называется серией в P, если

$p_{t+1} = p_{t+2} = ... = p_{t+v}$;

$p_t \ne p_{t+1}$; при $t \ge 1$;

$p_{t+v} \ne p_{t+v+1}$; при $t + v \le n$.

Число символов в серии называется длиной серии. i-ю серию будем описывать в виде $S_i(a_i,v_i)$, где a_i – символ, образующий i-ю серию, v_i – длина i-й серии. Символьную (r, n) последовательность P, состоящую из h серий будем называть (r, n, h) серийной последовательностью и представлять в виде

$$P = S_1(a_1,v_1)S_2(a_2,v_2)\ldots S_h(a_h, v_h), \quad (1)$$

где $a_i \in \Omega$;

$$\sum_{i=1}^{h} v_i = n .$$

Множество $A = \{a_1, a_2, \ldots , a_h\}$ называется структурой серийной последовательности, а множество $V = \{v_1, v_2, ..., v_h\}$ – составом серийной последовательности.

Серия $S_j(a_j, v_j)$ называется серией i-го вида, если $a_j = \omega_i$.

Количество серий i-го вида γ_i и количество символов i-го вида λ_i в последовательности P определяются следующим образом:

$$\gamma_i = \sum_{j=1}^{h} \beta_j , \quad (2)$$

$$\lambda_i = \sum_{j=1}^{h} v_i\beta_j , \quad (3)$$

где $$\beta_j = \begin{Bmatrix} 1, \text{если } a_j = \omega_i, \\ 0, \text{в противном случае} \end{Bmatrix}.$$

Множества $\Gamma=\{\gamma_1,\gamma_2,...,\gamma_r\}$ и $\Lambda=\{\lambda_1, \lambda_2,..., \lambda_r\}$ обладают следующими свойствами:

$$\sum_{i=1}^{r} \gamma_i = h, \qquad \sum_{i=1}^{r} \lambda_i = n.$$

На множестве символьных последовательностей введены следующие операции.

1. Слиянием последовательности $P^1 = S^1_1(a^1_1,v^1_1)\ldots S^1_{h1}(a^1_{h1}, v^1_{h1})$ с алфавитом Ω_1 и последовательности $P^2 = S^2_1(a^2_1,v^2_1)\ldots S^2_{h2}(a^2_{h2}, v^2_{h2})$ с алфавитом Ω_2 (обозначается $P^3 = \vartheta(P^1, P^2)$) называется последовательность $P^3 = S^3_1(a^3_1,v^3_1) \ldots S^3_{h3}(a^3_{h3},v^3_{h3})$ с алфавитом $\Omega_3 = \Omega_1 \cup \Omega_2$. Количество серий в последовательности P^3 зависит от вида символов a^1_{h1} и a^2_1, следующим образом.

При $a^1_{h1} \ne a^2_1$ последовательность P^3 имеет вид: $P^3 = S^1_1(a^1_1,v^1_1) \ldots S^1_h(a^1_h, v^1_{h1}) S^2_1(a^2_1,v^2_1) \ldots S^2_{h2}(a^2_{h2}, v^2_{h2})$, $h_3 = h_1 + h_2$.

При $a^1_h = a^2_1$ последовательность P^3 имеет вид: $P^3 = S^1_1(a^1_1,v^1_1) \ldots S^1_{h-1}(a^1_{h-1}, v^1_{h-1}) S^*(a^*,v^*) S^2_2(a^2_2,v^2_2) \ldots S^2_{h2}(a^2_{h2}, v^2_{h2})$,

где $a^* = a^1_h$, $v^* = v^1_{h1} + v^2_1$, $h_3 = h_1 + h_2 - 1$.

2. Выделение подпоследовательности в серийной последовательности $P^1 = S_1(a_1,v_1) S_2(a_2,v_2) \ldots S_{h1}(a_{h1}, v_{h1})$ (обозначается $P^2 = \Theta(P^1, \alpha, \beta)$), состоит в формировании последовательности P^2 вида $S_\alpha(a_\alpha, v_\alpha)\ldots S_\beta(a_\beta, v_\beta)$, т.е. выделение серий с номерами с α по β.

3. Вставка последовательности P^2 в последовательность P^1, начиная с μ-ой серии (обозначается $P^3 = \Psi(P^1, P^2, \mu)$) определяется следующим образом:

$P^3 = \Psi(P^1, P^2, \mu) = \vartheta(\Theta(P^1, 1, \mu - 1), P^2, \Theta(P^1, \mu, v^1_{h1}))$.

4. Соединением n - ичных последовательностей $P^1, P^2, \ldots , P^k$ (обозначается $P^\nabla = \Phi(P^1, P^2, \ldots , P^k)$) называется последовательность P^∇, элементы которой (слова) формируются следующим образом

$$p^\nabla_i = p^1_i p^2_i \ldots p^k_i,\ i = 1, \ldots , n.$$

Количество различных слов в последовательности P^∇ обозначим через ρ.

Определим количество (r, n) последовательностей, обозначенное L(r,n). Количество возможных r-ичных слов длиной n равно r^n. В рассматриваемых (r, n) последовательностях должны быть представлены все символы из алфавита Ω, поэтому среди множества возможных r-ичных слов необходимо исключить слова, содержащие не все символы из заданного алфавита, т.е.

$$L(r,n) = r^n - \sum_{i=1}^{r-1} \left(L(i,n) \times C_r^i \right), \quad (4)$$

где $L(1,n) = 1$.

В табл. 1 приведены значения L(r, n) для r = 2,...,6, n = 2,...,10, а в табл. 2 – количество (r, n) последовательностей по отношению ко всем n-ичным последовательностям.

Таблица 1

Количество (r, n) последовательностей для r = 2,...,6, n = 2,...,10.

n/r	2	3	4	5	6
2	2				
3	6	6			
4	14	36	24		
5	30	150	240	120	
6	62	540	1560	1800	720
7	126	1806	8400	16800	15120
8	254	5796	40824	126000	191520
9	510	18150	186480	834120	1905120
10	1022	55980	818520	5103000	16435440

Поскольку доля (r, n) последовательностей по отношению ко всем n-ичным последовательностям существенная, то для построения множества (r,n) последовательностей можно использовать последовательную генерацию n разрядных последовательностей $P = \{p_1,\ldots, p_n)$, i –ый элемент которой $p_i \in \Omega$ ($i = 1,\ldots, n$), определения количества различных символов в последовательности (δ) и отбор последовательностей с $\delta = r$. В таблице (табл. 3) приведены примеры (3,4) последовательностей.

Таблица 2

Количество (r, n) последовательностей по отношению ко всем n-ичным последовательностям для r = 2,...,6, n=2,...,10

n/r	2	3	4	5	6
2	50				
3	75	22,2			
4	87,5	44,4	9,38		
5	93,8	61,7	23,4	3,84	
6	96,9	74,1	38,1	11,5	1,54
7	98,4	82,6	51,3	21,5	5,40
8	99,2	88,3	62,3	32,3	11,4
9	99,6	92,2	71,1	42,7	18,9
10	99,8	94,8	78,1	52,3	27,2

Таблица 3

Каталог (3,4) последовательности

№	p_1	p_2	p_3	p_4	№	p_1	p_2	p_3	p_4
1	0	0	1	2	19	1	1	2	0
2	0	0	2	1	20	1	2	0	0
3	0	1	0	2	21	1	2	0	1
4	0	1	1	2	22	1	2	0	2
5	0	1	2	0	23	1	2	1	0
6	0	1	2	1	24	1	2	2	0
7	0	1	2	2	25	2	0	0	1
8	0	2	0	1	26	2	0	1	0
9	0	2	1	0	27	2	0	1	1
10	0	2	1	1	28	2	0	1	2
11	0	2	1	2	29	2	0	2	1
12	0	2	2	1	30	2	1	0	0
13	1	0	0	2	31	2	1	0	1
14	1	0	1	2	32	2	1	0	2
15	1	0	2	0	33	2	1	1	0
16	1	0	2	1	34	2	1	2	0
17	1	0	2	2	35	2	2	0	1
18	1	1	0	2	36	2	2	1	0

Для каждой (r, n) последовательности определяется ее структура (A), состав (V), номер типа серийной последовательности (T). Две серийные последовательности относятся к одному типу, если они имеют одинаковую структуру (V) и отличаются составом. В табл. 4 приведены характеристики (3, 4) последовательностей. Каталог типовых (3, 4) серийных последовательностей приведен в табл. 5.

Минимальное количество серий $h_{min} = r$, при $v_i = 1$, а максимальное количество серий $h_{max} = n$ при $v_i \geq 1$ ($i = 1,..., n$).

Множество типовых серийных (r, n) последовательностей $T(r, n)$ мощностью $G(r, n)$ состоит из множества типовых серийных последовательностей $T(r, n, h)$ мощностью $G(r, n, h)$ для $h = r, \ldots, n$.

Таблица 4

Характеристики (3,4) последовательностей

№ п/п	A				V				Te
	a_1	a_2	a_3	a_4	v_1	v_2	v_3	v_4	
1	0	1	2		2	1	1		1
2	0	2	1		2	1	1		2
3	0	1	0	2	1	1	1	1	3
4	0	1	2		1	2	1		1
5	0	1	2	0	1	1	1	1	4
6	0	1	2	1	1	1	1	1	5
7	0	1	2		1	1	2		1
8	0	2	0	1	1	1	1	1	6
9	0	2	1	0	1	1	1	1	7
10	0	2	1		1	1	2		2
11	0	2	1	2	1	1	1	1	8
12	0	2	1		1	2	1		2
13	1	0	2		1	2	1		9
14	1	0	1	2	1	1	1	1	10
15	1	0	2	0	1	1	1	1	11
16	1	0	2	1	1	1	1	1	12
17	1	0	2		1	1	2		9
18	1	0	2		2	1	1		9
19	1	2	0		2	1	1		13
20	1	2	0		1	1	2		13
21	1	2	0	1	1	1	1	1	14
22	1	2	0	2	1	1	1	1	15
23	1	2	1	0	1	1	1	1	16
24	1	2	0		1	2	1		13
25	2	0	1		1	2	1		17
26	2	0	1	0	1	1	1	1	18
27	2	0	1		1	1	2		17
28	2	0	1	2	1	1	1	1	19
29	2	0	2	1	1	1	1	1	20
30	2	1	0		1	1	2		21
31	2	1	0	1	1	1	1	1	22
32	2	1	0	2	1	1	1	1	23
33	2	1	0		1	2	1		21
34	2	1	2	0	1	1	1	1	24
35	2	0	1		2	1	1		17
36	2	1	0		2	1	1		21

Таблица 5

Каталог типовых (3, 4) серийных последовательностей

№	a_1	a_2	a_3	a_4	Кол.	№	a_1	a_2	a_3	a_4	Кол.
1	0	1	2		3	13	1	0	1	2	1
2	0	2	1		3	14	1	0	2	0	1
3	1	0	2		3	15	1	0	2	1	1
4	2	0	1		3	16	1	2	0	1	1
5	1	2	0		3	17	1	2	0	2	1
6	2	1	0		3	18	1	2	1	0	1
7	0	1	0	2	1	19	2	0	1	0	1
8	0	1	2	0	1	20	2	0	1	2	1
9	0	1	2	1	1	21	2	0	2	1	1
10	0	2	0	1	1	22	2	1	0	1	1
11	0	2	1	0	1	23	2	1	0	2	1
12	0	2	1	2	1	24	2	1	2	0	1

Значения G(r, n) и T(r, n) определяются следующим образом:

$$G(r,n)=\sum_{h=r}^{n}G(r,n,h)\,,\quad T(r,n)=\bigcup_{h=1}^{n}T(r,n,h).$$

Построение множества T(r, n) состоит из следующих этапов:

Этап 1. Полагаем h = r.

Этап 2. Формируем множество T(r, n, r).

Этап 3. Полагаем h = h+1.

Этап 4. Если h>n, то переходим к этапу 7.

Этап 5. На основе множества T(r, n, h-1) формируем множество T(r, n, h).

Этап 6. Переходим к этапу 3.

Этап 7. Конец

Рассмотрим формирование множества T(r, n, r) на втором этапе.

Очевидно, что G(r, n, r) = r и множество типовых серийных (r, n, r) последовательностей T(r, n, r) получается в результате перестановок элементов последовательности $P = \{p_1,\ldots, p_r\}$.

На пятом этапе на основе множества T(r, n, h-1) формируется множество T(r, n, h) следующим образом.

Для каждой типовой последовательности $P^T_i \in T(r, n, h\text{-}1)$, i=1,..., G(r, n, h-1) с алфавитом Ω формируется множество последовательностей E, полученных в результате добавления в последовательность P^T_i последовательности $P^* = S^*(a^*,1)$, где $a^* \in \Omega$.

В состав множества последовательностей E входят последовательности, полученные с помощью следующих операций:

1. $P^E = \vartheta(P^*, P^T_i)$.
2. $P^E = \Psi(P^T, P^*, \mu)$, μ=2, ..., h – 1.
3. $P^E = \vartheta(P^T_i, P^*)$.

Для всех последовательностей, входящих во множество E определяется тип серийной последовательности. Множество различных типовых представителей образуют множество типовых серийных последовательностей T(r, n, h). В табл. 6 приведены значения G(r, n, h) для r = 2,..., 6, h = r, ... , 9, n ≥ h.

Таблица 6

Значения G(r, n, h) для r = 2,..., 6, h = r, ... , 9, n ≥ h

r\h	2	3	4	5	6	7	8	9
2	2	2	2	2	2	2	2	2
3		6	18	42	90	186	378	762
4			24	144	600	2160	7224	23184
5				120	1200	7800	42000	204120
6					720	10800	100800	756000

В табл. 7 приведены значения G(r, n) для r = 2,..., 6, n = r, ... , 9, n ≥ h ≥ r.

Таблица 7

Значения G(r, n) для r = 2,..., 6, n = r, ... , 9, n ≥ h ≥ r

r\n	3	4	5	6	7	8	9
2	4	6	8	10	12	14	16
3	6	24	66	156	342	720	1482
4		24	168	768	2928	10152	33336
5			120	1320	9120	13320	217440
6				720	11520	112320	868320

Количество (r, n) последовательностей W, с типовой структурой $A = \{a_1, a_2, \ldots , a_h\}$, множествами $\Gamma=\{\gamma_1,\gamma_2,\ldots,\gamma_r\}$ и $\Lambda=\{\lambda_1, \lambda_2,\ldots, \lambda_r\}$ определяется количеством композиций λ_i символов на γ_i серий i = 1,...,r:

$$W=\prod_{i=1}^{r}\delta(\lambda_i,\gamma_i),$$

где $\delta(t,d)=\dfrac{(t-1)!}{(d-1)!(t-d)!}$.

Например, для $A = \{a_1, a_2, a_1, a_2, a_3, a_2, a_3\}$, Γ = {2, 3, 2} и Λ = {5, 4, 3} существует 24 варианта (3, 12) последовательностей, состав которых приведен в табл. 8.

Таблица 8

Состав последовательностей

№	v_1	v_2	v_3	v_4	v_5	v_6	v_7	№	v_1	v_2	v_3	v_4	v_5	v_6	v_7
1	4	2	1	1	2	1	1	13	3	2	2	1	2	1	1
2	4	2	1	1	1	1	2	14	3	2	2	1	1	1	2
3	4	1	1	2	2	1	1	15	3	1	2	2	2	1	1
4	4	1	1	2	1	1	2	16	3	1	2	2	1	1	2
5	4	1	1	1	2	2	1	17	3	1	2	1	2	2	1
6	4	1	1	1	1	2	2	18	3	1	2	1	1	2	2
7	1	2	4	1	2	1	1	19	2	2	3	1	2	1	1
8	1	2	4	1	1	1	2	20	2	2	3	1	1	1	2
9	1	1	4	2	2	1	1	21	2	1	3	2	2	1	1
10	1	1	4	2	1	1	2	22	2	1	3	2	1	1	2
11	1	1	4	1	2	2	1	23	2	1	3	1	2	2	1
12	1	1	4	1	1	2	2	24	2	1	3	1	1	2	2

Заключение

Предложенный метод перечисления символьных последовательностей, основанный на представлении их в виде серийных последовательностей, позволяет создавать каталоги типовых символьных последовательностей с заданными свойствами. Дальнейшее направление исследований – разработка алгоритмического и программного обеспечения для автоматизации построения каталогов серийных последовательностей.

Литература

1. Roytberg, M.A. A Search for Common Patterns in many Sequences [Text] /M.A. Roytberg // Comput. Appl. Biosci. – 1992. – Vol. 8, № 1. – P. 57 – 64.

2. Braun, V. An Enumerative Coding Technique for DC-Free Runlength-Limited Sequences [Text]/V. Braun, K.S. Immink // IEEE Trans. Commun. – 2000. – V. 48, №. 12. – P. 2024-2031.

3. Курмаев, О.Ф. Кодирование последовательностей с ограниченными длинами серий [Текст]/ О.Ф. Курмаев // Пробл. передачи информ. – 2001. – Т. 37, № 3. – С. 34-43.

4. Амелькин, В.А. Пересчет, нумерация и генерирование серийных последовательностей с отделенными натуральными сериями [Текст] /В.А. Амелькин // Сиб. журн. вычисл. математики / РАН. Сиб. отд-ние. – Новосибирск, 2006. – Т. 9, № 2. – С. 109–121.

5. Амелькин, В.А. Нумерация неубывающих и невозрастающих серийных последовательностей [Текст] /В.А. Амелькин // Сиб. журн. вычисл. математики / РАН. Сиб. отд-ние. – Новосибирск, 2009. – Т. 12, № 4. – С. 389–401.

6. Golic, J.Dj. Constrained embedding probability for two binary strings [Text]/ J.Dj. Golic // SIAM J. Discrete Math. – 1996. – V.9, No. 3. – P. 360-364.

Поступила в редакцию 24.09.2012

Рецензент: д-р техн. наук, проф., проф. кафедры информатики М.Л. Угрюмов, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков.

МЕТОД ПЕРЕЛІКУ СИМВОЛЬНИХ ПОСЛІДОВНОСТЕЙ

М.Д. Кошовий, О.М. Костенко, Н.В. Доценко, Г.В. Павлик

Розглянуто властивості символьних послідовностей і способи їхнього формального подання у вигляді серійних послідовностей. Введені операції на множині символьних послідовностей, що дозволяють проводити їх перетворення. Досліджено способи опису серійних послідовностей і їхніх типових представників. Запропоновано метод перерахування символьних послідовностей, що дозволяє створювати каталоги типових символьних послідовностей із заданими властивостями. Отримано оцінки кількості типових серійних послідовностей і процедура формування множини типових варіантів. Розглянуто приклади реалізації запропонованого методу.

Ключові слова: символьна послідовність, серія, перерахування, типовий представник.

METHOD OF SYMBOLICAL SEQUENCES ENUMERATION

N.D. Koshevoj, E.M. Kostenko, N.V. Dotsenko, A.V. Pavlik

Properties of symbolical sequences and ways of their formal representation in the form of serial sequences are considered. Operations on set of the symbolical sequences are entered, allowing spends their transformations. Ways of the serial sequences description and their typical representatives are investigated. The method of the symbolical sequences enumeration, allowing to create catalogues of typical symbolical sequences with the set properties is offered. Estimations of typical serial sequences quantity and procedure of typical variants set formation are received. Examples of the offered method realization are considered.

Keywords: symbolical sequence, a series, transfer, the typical representative.

Кошевой Николай Дмитриевич – д-р техн. наук, проф., заведующий кафедрой авиационных приборов и измерений, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков, Украина, e-mail: kafedraapi@rambler.ru.

Костенко Елена Михайловна – канд. техн. наук, доцент Полтавской Государственной аграрной академии, Полтава, Украина, e-mail: Kostenko@pdaa.com.ua.

Доценко Наталья Владимировна – канд. техн. наук, доцент, доцент кафедры менеджмента, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков, Украина, e-mail: dotsenko2006@ukr.net.

Павлик Анна Владимировна – ассистент кафедры авиационных приборов и измерений, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков, Украина, e-mail: pavlan2@ukr.net.